Article

Journal of Early Childhood Literacy
2020, Vol. 20(2) 223–251

DOI: 10.1177/1468798417753713
journals.sagepub.com/home/ecl
SAGE

Measuring young children's word knowledge: A conceptual review

Elizabeth Burke Hadley
University of South Florida, USA

David K Dickinson
Vanderbilt University, USA

Abstract
The importance of early vocabulary development to later reading comprehension has been well-established. However, there have been a number of criticisms that the assessments typically used to measure oral vocabulary knowledge do not adequately capture the complexity of this construct. This conceptual review works towards a more robust theoretical framework for vocabulary knowledge, focusing especially on the understudied dimension of vocabulary depth, which can be used to evaluate and design measures for early childhood learners. This framework is then used to review measures commonly used for preschool to 1st grade learners in the context of vocabulary interventions and observational studies. Recommendations are made for the use of existing measures and the design of future measures.

Keywords
Assessment, depth, language development, pre-school children, vocabulary

To assess word knowledge, one must first determine what it means to know a word. Many assessments used for early childhood learners count a word as "known" if a child can identify an image representing a word. These known word items can then be tallied to give a rough estimate of a child's vocabulary size, often referred to as breadth of knowledge. A key question emerges from this emphasis on vocabulary breadth, however: what is the *quality* of learners'

Corresponding author:
Elizabeth Burke Hadley, University of South Florida, Department of Teaching and Learning, 4202 E. Fowler Ave., Tampa, FL 33620, USA.
Email: hadleye@usf.edu

knowledge of these ''known'' words, and how might further investigating this aspect of vocabulary enrich our understanding of children's word learning?

We argue here that early childhood vocabulary assessment would benefit from a richer, more multifaceted theoretical foundation that takes into account the complexity of what it means to know a word. We focus specifically on the understudied dimension of vocabulary depth, or the quality of one's word knowledge. For example, a child might be able to correctly select a picture that depicts a target word, but such an assessment leaves open the question of what the child is really able to *do* with that word. Are they also able to pronounce that word correctly, use it in a sentence or understand its meaning in a text? Do they know the situations in which this word typically would or would not be used? Given that these are the functional vocabulary skills necessary for proficient reading, writing and speaking, vocabulary depth must be studied if we are to better support young children's language and literacy development.

We have called this paper a conceptual review in order to distinguish it from systematic reviews, which typically ask a specific empirical question phrased in a cause-and-effect form (e.g. ''To what extent does A contribute to B?'') (Kennedy, 2007: 139). In contrast, a conceptual review is intended to gain new insights into a problem (Kennedy, 2007) and to ''map the terrain'' of an area of research to spur further enquiry (Kucan and Palincsar, 2010: 341). Here, we wish to lay the groundwork for a more robust theoretical framework that offers insights into the evaluation and design of vocabulary measures for young children. Using this framework, we map the terrain of early childhood vocabulary assessments, reviewing the measures most commonly used with preschool to 1st grade students. We end by making recommendations for the use and design of measures.

Vocabulary assessment in early childhood

Recent articles (Hoffman et al., 2014; Pearson et al., 2007) have expressed concerns about the impoverished state of vocabulary assessment. Pearson et al. (2007), describing primarily print-based measures, characterized this field as ''grossly undernourished, both in its theoretical and practical aspects'', and called for measures that ''are as conceptually rich as the phenomenon (vocabulary knowledge) they are intended to measure'' (2007: 282–283). Pearson et al. (2007) further argued that future research should focus on distinguishing between different dimensions of vocabulary knowledge, rather than using vocabulary as a global term.

Hoffman et al. (2014) were similarly critical of the current state of assessment. Like the present paper, their paper examined the challenge of assessing oral vocabulary in young children and identified breadth and depth as important dimensions to consider for assessment. Hoffman et al. (2014) focused mainly on the difficulty of assessing word-learning from book-reading interventions and based their criticism of measures on their lack of reliability and validity. They concluded that many measures are convenient, but not practically significant or well-designed.

The present paper builds on both Pearson et al.'s (2007) and Hoffman et al. (2014)'s work. We see this paper as responding to Pearson et al.'s call (2007) to identify more conceptually rich measures and to highlight additional dimensions of vocabulary knowledge, but we focus here on the unique challenge of assessing oral language in an early childhood population. We build on and complement Hoffman et al.'s (2014) review of oral vocabulary assessments by looking at a wider range of interventions, including correlational work and evaluating measures based on a theoretical framework for word-learning rather than their psychometric properties.

Both Pearson et al. (2007) and Hoffman et al. (2014) characterize vocabulary assessment as undernourished and inadequate. Early childhood vocabulary assessment is particularly susceptible to such criticism, as many commonly used vocabulary measures rely on breadth as the main indicator of word knowledge. The pitfalls of this approach can be illustrated through the findings of the National Early Literacy Panel's (NELP, 2008) meta-analysis of relationships between early abilities and later conventional literacy skills. Many were surprised that the NELP report found that early (preK-K) vocabulary was a relatively weak predictor of reading comprehension in kindergarten-2nd grade ($r = 0.25$) (Dickinson et al., 2010; NELP, 2008). The definition of vocabulary used in the report was quite narrow, however, encompassing only vocabulary breadth (as measured by receptive assessments). Vocabulary depth (as measured by giving definitions of words), on the other hand, was a significantly stronger predictor of reading comprehension ($r = 0.45$), along with other complex oral language skills such as grammar and listening comprehension. Additional research corroborates the NELP report's (2008) findings: depth has been found to play a unique and important role in the relationship between vocabulary and reading comprehension (Roth et al., 2002), even when controlling for breadth (Cain and Oakhill, 2014; Ouellette, 2006; Proctor et al., 2012). While using breadth measures to assess vocabulary in early childhood studies is the norm, these results indicate that this reliance on breadth tells an incomplete story about the relationship between vocabulary and reading comprehension.

Further evidence suggests that the measures used to capture growth in vocabulary knowledge are similarly inadequate. In a vocabulary intervention meta-analysis performed by Marulis and Neuman (2010), nearly half of the studies included used only standardized assessments, even though such measures are insensitive to small increases in knowledge (National Reading Panel, 2000). Furthermore, nearly all of the standardized measures used were picture vocabulary measures, such as the Peabody Picture Vocabulary Test (PPVT), which are intended to estimate breadth, not depth. Of the interventions that used a researcher-created measure to track the learning of specific words, only about half used a measure that tapped into depth. This focus on breadth probably misses information about children's word-learning that could better inform instructional approaches.

Dimensions of vocabulary knowledge: Breadth and depth

The following section defines the terms breadth and depth separately. Our intention here, however, is not to suggest that breadth and depth are two completely separate constructs. Rather, they are interrelated aspects of vocabulary knowledge that probably grow in tandem (e.g. Tannenbaum et al., 2006). As a word-learner adds new words to her vocabulary, she also gains more knowledge about how those words relate to one another, thus adding to the overall quality of her word knowledge. Given that these aspects are related, but offer different information about word-learning, we argue here that measuring both dimensions of word knowledge can lead to a richer, more detailed portrait of word-learning.

Defining breadth

Vocabulary breadth is an estimate of the overall number of items in one's lexicon, without specific attention being paid to how well each item is known. The term "breadth" is often used to denote a shallower, less comprehensive understanding of individual words (e.g. Hoffman et al., 2014), such as "fast-mapped" knowledge, an initial representation of a word gained through only a few exposures that consist mainly of phonological or syntactic information (Carey, 1978).

The breadth of young children's vocabularies has been shown to be predictive of their language and literacy achievement well into elementary school (Lee, 2011; Senechal et al., 2006; but see NELP, 2008, which found that this predictive power is relatively weak in the early grades). Important research has also demonstrated that vocabulary size varies by socioeconomic status (SES)

(Qi et al., 2006). As these studies demonstrate, measuring vocabulary in terms of numbers of words known has been a fruitful direction, and one that has dominated the field. This perspective on vocabulary is necessary but, as we argue here, not sufficient.

Defining depth

There has been much confusion in the literature around the concept of vocabulary depth (Schmitt, 2014). The classic, highly influential definition of depth comes from Anderson and Freebody's work on the relationship between vocabulary and reading comprehension, which defines breadth as how many words are known and depth as how well those words are known (1981). This definition is powerful in terms of its simplicity and clarity, and it has often been repeated in the literature (e.g. Read, 2004). When it comes to operationalizing depth, however, this definition is somewhat empty. What exactly does it mean to know a word well or less well?

Depth has also been defined as a learner's knowledge of multiple aspects of a word, including its phonological, orthographic, pragmatic, syntactic and semantic features (Silverman and Hartranft, 2015). This particular conception of depth has been described as a "components" approach, one that looks at word knowledge as a collection of different elements (Schmitt, 2014). The components approach has been influential, especially in the literature on second language learners (Schmitt, 2014), and it adds much-needed specificity to Anderson and Freebody's (1981) definition. However, the components approach becomes somewhat unwieldy as various definitions of depth include not only the aspects mentioned above, but also features such as morphology, collocations and grammatical functions (Nation, 2001). A more cohesive idea of depth is needed that has a strong theoretical position about how these elements relate to one another and what constitutes "deeper" knowledge for each one. A components approach also lacks a theory of why depth might contribute to better reading comprehension and how deep word-learning progresses over time. Therefore, we draw on a more robust, theoretically rich conceptualization of depth for this paper that addresses many of these concerns: Perfetti's (2007) Lexical Quality Hypothesis (LQH).

Depth as lexical quality. Perfetti's LQH (2007) views reading comprehension as dependent on the ability to efficiently retrieve word identities. This ability is in turn based on the lexical quality of a word, or how much knowledge a child has about the form and meaning of a particular word, as well as how closely

these form and meaning elements are connected to one another. The LQH presents depth as a continuum of knowledge, with individual words ranging from low to high lexical quality. High-quality representations are those that have stable representations of spelling and sound and a generalized sense of word meaning that can transfer to multiple contexts. These elements are also closely connected, making the retrieval of a complete word identity automatic. In contrast, low-quality lexical representations are those in which the phonological representation is less stable and meaning may be bound to a single context. The elements of form and meaning are loosely connected, causing inefficient or inaccurate retrieval of a word's identity (e.g. hearing a word and mistakenly retrieving the meaning of a similar-sounding word). Therefore, low quality representations threaten a reader's ability to comprehend a passage.

The LQH is built on an understanding of the word learning process as incremental, with each additional encounter with a word improving the quality of its representation. This incremental process of word-learning generally begins with a quick, relatively crude mapping of a word to an object or action (Hollich et al., 2000). Research on early language development shows that children can "fast map" (Carey, 1978), or gain a shallow understanding of a word, thus adding it to their lexicon, without fully understanding the multiple facets of its meaning (Lahey, 1988). Indeed, 1.5–2 year-olds are able to establish very minimal representations of word forms, consisting mainly of phonological (Estes et al., 2007; Swingley, 2007) or syntactic (Yuan and Fisher, 2009) representations of words that can be filled in later with more comprehensive semantic information. These initial, shallow representations of words are quite fragile (Bion et al., 2013) and may be forgotten if not reinforced (Horst and Samuelson, 2008). Children who have fast-mapped a word may be able to successfully identify its image on a receptive vocabulary measure, but they lack deeper conceptual and pragmatic knowledge of the word. Therefore, while in some sense children may "know" fast-mapped words, these low quality lexical representations may not support real-world comprehension or use.

How, then, do learners acquire high quality word knowledge? Proceeding further along the continuum to higher-quality representations of words takes time and more encounters with words (Bion et al., 2013; Yu and Smith, 2012), a process sometimes described as "slow-mapping" (Swingley, 2010). The Instance-Based Learning Framework (Bolger et al., 2008), provides an incremental model for depth of word learning in which each encounter with a word encodes a trace of its form and context in memory. Frishkoff et al. (2011)

measured adults' incremental learning of new words, finding that the definitions they gave increased in accuracy with each additional word-learning episode. Encountering a word in multiple varied contexts, and the provision of explicit definitions for words, further expedites the building of high quality representations (Bolger et al., 2008; Yurovsky et al., 2014).

The work discussed so far provides a useful theoretical framework for deep word learning, but it has been primarily focused on reading and therefore mainly validated with older children and adults. There are few available studies about how preschool-aged children acquire higher-quality word knowledge (but see Hadley et al., 2016). However, the LQH is equally relevant to early childhood learners, in that it highlights the importance of building high quality word meanings, a protracted, incremental process that must begin early if later reading comprehension is to proceed smoothly. To apply the LQH to these young learners requires an understanding of the term "form" that refers to its phonological and grammatical aspects, rather than the orthographic features that are emphasized in most discussions of lexical quality (Perfetti, 2013; Verhoeven and Perfetti, 2011).

Several researchers have suggested that being able to use a word in an appropriate way is another important aspect of lexical quality (Nagy and Scott, 2000; Read, 2004; Silverman and Hartranft, 2015). Use is not one of the main constituents of depth in the LQH, although Perfetti sees the quick retrieval of a word's identity (i.e. the ability to remember or use a word in context) as the key indicator of high quality word knowledge. Defining lexical quality in terms of "use" is even more important when studying young children, whose knowledge of words is particularly tied to context (Snow et al., 1991).

Therefore, we define depth at the word level here as the richness of one's representations of a word, including grammatical, phonological and semantic information, the ability to use a word, and how tightly bound these elements are to one another (see Ordóñez et al., 2002). In other words, lexical quality is comprised of form, meaning and use (Nation, 2001). We consider these constituents and their development separately below, with the understanding that, in practice, they are closely bound to one another.

Form. The first important constituent of lexical quality is that of form. For young children who are not yet reading print, form refers to the phonological representation of a word, or the storage of information about the sounds in a particular word (Ainsworth et al., 2015). Like other aspects of lexical quality, the quality of a phonological representation can range from low to high. Quality is determined by the accuracy and stability of the representation.

Another aspect of form includes the grammatical features of a word. According to Perfetti (2007), high quality knowledge of a word's form means knowing all grammatical classes of the word (e.g. knowing both *anger* and *angry*) and being able to manipulate the word to reflect changes in tense, mood (e.g. the conditional mood: I *would have* eaten there), person, number and gender. A lower-quality representation of a word might mean that the learner was unable to use inflected forms of a word consistently and appropriately.

Meaning. The second constituent of lexical quality to be considered is that of meaning. In Perfetti's LQH (2007), having higher-quality semantic information about a word allows a person to distinguish between closely related words. For example, in order to discriminate between a *shovel* and a *spade*, it is necessary to understand what each is used for, their relative sizes and other perceptual features of each.

The kind of semantic information it is possible to learn about words varies by word type, with different kinds of semantic information available for highly imageable, concrete words than for more abstract words (Hadley et al., 2016). Concrete nouns name parts of the world that are naturally individuated, whereas most verbs, adjectives and abstract nouns label more diffuse, relational concepts such as ideas or qualities (Gentner, 2006). Therefore, the kind of meaning information that can be learned about these words is qualitatively different.

In particular, concrete nouns have a much wider array of semantic information categories available than do other word types because of their perceptual accessibility. Perceptual information is sometimes considered to be a gateway to a deeper conceptual understanding of words (Booth and Waxman, 2002), leading to the kind of refined, precise knowledge that is characteristic of high lexical quality. For example, feeling that a *helmet* is hard helps children to understand its function (to protect someone's head), while seeing that a creature has four legs helps a child to categorize that object as an *animal*. These examples illustrate two of the types of conceptual information potentially available for concrete nouns: functional information, what something does or is used for (e.g. a *helmet* protects one's head), and category membership (a cat is a kind of *animal*). Of course, being able to identify a word as an instance of a category is not something that is supplied simply by seeing an object; rather, that concept must have been previously constructed. This brings into focus another dimension of semantic depth: the extent to which words reside in rich semantic networks, with some research showing that young

children may be more likely to learn new words that are semantically related to known words than those that are unrelated (Borovsky et al., 2016).

Other types of semantic information apply both to concrete nouns and more abstract words. For example, synonyms, or a core meaning for a word, are available for many words across word types (Miller and Fellbaum, 1991). There are also some semantic features that are unique to word types other than concrete nouns. For example, for three-year-olds learning action verbs, causation, or who or what caused the action, is the most salient feature (Forbes and Farrar, 1993).

Use. Use refers to the ability to put word knowledge into action, such as appropriately using a word in multiple contexts to convey meaning (Silverman and Hartranft, 2015), as well as awareness of a word's connotations, typical registers, idiomatic or rhetorical uses. This aspect of word knowledge is sometimes seen as the true marker of high quality word knowledge, where "knowing a word means being able to do things with it" (Nagy and Scott, 2000: 237). Under the LQH, the rapid retrieval and use of a word is also the hallmark of high quality semantic and phonological knowledge.

The ability to appropriately *use* a word develops slowly over time, progressing from comprehension of a word, to use of a word in a single context, to eventually using the word across a range of contexts (Clark, 2010). Children often have surprisingly restricted contexts of use for words that they appear to "know". Seston et al. (2009) found that half of six-year-olds tested were unable to understand common verbs used in unfamiliar contexts (e.g. someone *sweeping* dirt away with their feet), even when the new context included a number of details to help support comprehension. These results show that while young children may "know" common verbs, they require further exposure and support to understand and use these words in proficient, flexible ways across a range of contexts.

Low quality knowledge of use includes a range of situations: when a child has only a memory trace of a word used in a certain context (Bolger et al., 2008), but is not yet able to use the word herself, or a child who has memorized a dictionary definition of a word, but whose actual use of that word is odd and/or incorrect. As knowledge of use develops, a child may be able to use a word correctly in a single context. High-quality use would include being able to use the word in several contexts, as well as generalizing its use to new contexts.

Summary. We review these aspects of form, meaning and use to indicate the range of quality that is possible for individual word items along each of these

dimensions. For purposes of assessment, this perspective demonstrates that knowing a word is not an all-or-nothing proposition, but rather can range from low to high in a variety of ways.

Review of selected vocabulary measures

In the following section, we will apply our theoretical framework to measures currently used in the field. We address the following question: how much, and what kinds, of knowledge do measures assess? With this question in mind, we review the major vocabulary assessments currently in use with preschool to 1st grade students in the context of vocabulary interventions or studies on the relationship between oral language development and reading comprehension. We do not include assessments used exclusively in clinical settings for screening of language difficulties (see Table 1 for a summary of the assessments reviewed).

How much, and what kind, of knowledge is assessed

We organize our review of measures into two major categories, based on the theoretical distinction made earlier between **breadth** and **depth**. The first set of measures to be discussed are **breadth** measures, assessments intended to gauge the number of words known by a learner, without particular attention paid to the quality of that knowledge. **Breadth** measures typically tap into shallower, lower lexical-quality knowledge of words (such as phonological and perceptual knowledge, see Table 1), in order to quickly yield a rough count of words "known", at least partially.

In contrast, **depth** measures are intended to assess lexical quality, and therefore tap into multiple aspects of word knowledge (see Table 1). Capturing the multiple facets of depth has proved difficult from an assessment perspective, as it is impractical to intensively assess the quality of each phonological, syntactic, semantic and pragmatic representation of words. Instead, most assessments attempt to capture two or three aspects of depth, reasoning that more advanced knowledge in one category may also demonstrate basic knowledge in the others (Read, 2004). This approach can still be quite time-consuming, which often means in practice that **depth** measures assess a smaller number of words more intensively than do **breadth** measures.

We address our guiding question by highlighting which aspects of word knowledge each measure taps into (form, meaning and use), and to what extent

Table 1. Breadth and depth vocabulary assessments in use with preschool-1st grade learners.

				Aspects of word knowledge assessed							
				Form		Meaning		Use in context			
Type of measure	Receptive/ expressive	Name of measure and ages	Standardized (S) or targeted (T)	Phonology	Grammar	Perceptual	Conceptual	Familiar	Novel	Ease of admin.	Ease of scoring
Breadth	Receptive	PPVT (2:6+)	S	X		X				E	E
	Receptive	PPVT-like formats	T	X		X					
	Expressive	EOWPVT (2:0+)	S	X		X				E	E
	Expressive	EOWPVT-like formats	T	X		X					
	Expressive	EVT (2:6+)	S	X		X				E	E
	Expressive	EVT-like formats	T	X		X					
	Both	WJ-III Picture Vocabulary subtest (2:0+)	S	X		X				E	E
Depth	Expressive	Definition tasks (targeted)	T	X	X	X	X			M	H
	Expressive	TOLD-P:4, Oral Vocabulary subtest (4:0–8:11)	S	X		X	X			M	M
	Expressive	TOWK, Word Definitions subtest (5–17)	S	X		X	X			M	M
	Expressive	CELF (formulated sentences) (5–21)	S	X	X	X	X		X	M	M
	Receptive	Closed-ended questions	T	X		X	X		X	E	E
	Expressive	Story-retelling and description	T	X	X	X	X	X[a]	X[b]	M	H
	Expressive	Context integration	T	X	X				X	E	M

PPVT: Peabody Picture Vocabulary Test; EOWPVT: Expressive One-Word Picture Vocabulary Test; EVT: Expressive Vocabulary Test; TOWK: Test of Word Knowledge; CELF: Clinical Evaluation of Language Fundamentals; E: easy; M: medium; H: hard.

[a]Penno et al. (2002).

[b]McKeown and Beck (2014).

(see Table 1). Form is divided into the two major aspects discussed earlier: (1) phonology (are children asked to recognize or produce an accurate phonological representation?) and (2) grammar (are children asked to recognize or produce inflections of a word?). The meaning category is comprised of (1) perceptual information, measures that tap into knowledge of a word's perceptual features and (2) conceptual information, measures that tap into knowledge of conceptual information, such as function and category membership. The use in context category includes (1) familiar, or measures that only test children's understanding or production of a word in a familiar context, and (2) novel, or measures that test children's understanding or production of a word in a new context. Note that the categories of form, meaning and use are not mutually exclusive – a single measure could theoretically test every aspect listed above.

We also categorize measures in terms of whether they assess receptive or expressive knowledge (see Table 1). Receptive measures, which test the understanding of words, are generally less difficult than expressive measures, which test the ability to correctly retrieve a word and produce its label (Melka, 1997). Receptive measures are often also **breadth** measures, as the format is well-suited to assessing a large number of words. Expressive measures vary in their difficulty: some require only that children produce the target word, while others ask them to use the word in a sentence. Some, but not all, expressive measures are also **depth** measures, depending on the goal of the assessment (whether they aim to measure quantity or quality).

The next column in Table 1 notes whether a measure is **standardized** or **targeted** (targeted measures are also referred to elsewhere as author-created, Marulis and Neuman, 2010, or researcher-created, e.g. Kieffer and Lesaux, 2012). **Standardized** measures are those which are intended to assess either the overall size or quality of a child's lexicon. In contrast, **targeted** measures assess a specific sample of words from a curriculum or intervention to track growth in learning.

Finally, we include columns on ease of administration and scoring. These columns are intended to address some of the real-world concerns involved in assessing children's word knowledge, and to acknowledge some of the trade-offs that must be made in terms of comprehensiveness of assessments vs time and energy.

Breadth measures

Peabody Picture Vocabulary Test and measures modelled after the PPVT. The PPVT (Dunn and Dunn, 2007) is perhaps the most widely used vocabulary measure for

young children. It is a standardized measure, normed on a national sample so that the results can be used to determine how a child's vocabulary compares to their peers. In administering the PPVT, the examiner reads a word out loud and asks the child to point to the referent from a choice of four illustrations. Test administration stops when a child makes eight or more errors in a set of 12 items.

The PPVT has been used most commonly in the literature in the following ways: (1) to determine whether the size of learners' vocabularies is within typical range for their age (e.g. Farkas and Beron, 2004); (2) in correlational studies that explore the relationship between early vocabulary and its impact on later reading achievement (e.g. Cunningham and Stanovich, 1997; Dickinson and Porche, 2011); and (3) to assess long-term or general growth in vocabulary knowledge – for example, to investigate how school instruction has impacted on vocabulary during a school year (e.g. Silverman and Crandell, 2010). The PPVT has also been used to measure whether a short-term vocabulary intervention affected general vocabulary growth (e.g. Hargrave and Sénéchal, 2000). However, there are some concerns that the PPVT may not be well-suited to that purpose, as teaching a small number of new words is unlikely to change a child's general vocabulary knowledge (National Reading Panel, 2000).

Due to concerns about using standardized measures such as the PPVT to track small increases in vocabulary knowledge, many short-term vocabulary interventions have created versions of the PPVT to track growth in learning for a specific set of words (we refer to these here as "PPVT-like measures"). An individual item in such measures typically includes a visual representation of a target word taught in the intervention, along with 3–4 foils. As in the PPVT, children are asked to point at the picture corresponding to the word read aloud by the examiner. These PPVT-like measures are widely used in early childhood vocabulary interventions to assess children's learning of target words (e.g. Pollard-Durodola et al., 2011; Roskos et al., 2008; Sénéchal, 1997; Wasik and Bond, 2001).

We apply our guiding question of "how much and what kind of knowledge is measured" to the PPVT and PPVT-like measures to track the extent to which they tap into form, meaning and use.

Form. Both the PPVT and PPVT-like measures tap into receptive knowledge of the phonological form of a word – a child must recognize the word as said by the examiner, then map it to a visual representation. Neither assesses grammatical inflections of a word.

Meaning. Both the PPVT and PPVT-like measures tap into semantic knowledge at the shallow end of the lexical quality continuum. In general, to get an item "correct", a child must have mapped a word's perceptual features to its label, but may or they may not have deeper conceptual knowledge about the word. However, some of the most difficult items on the PPVT (those intended for older children and adults, not typically encountered by young children) do test more finely differentiated conceptual knowledge by including foils that are very closely related to the target item. In contrast, PPVT-like measures typically do not include progressively harder items, since each item tests knowledge of a word taught in the intervention. Items are therefore intended to be of roughly equivalent difficulty.

Tapping into shallower semantic knowledge has several advantages: these measures can detect fairly minimal amounts of knowledge, which can be helpful when assessing children who have very low levels of vocabulary knowledge or language skills, or who are reluctant to speak to an examiner. The standardized PPVT is a highly efficient method of determining the size of a child's lexicon and can be used to track large increases in generalized word knowledge. PPVT-like measures are similarly efficient and can be used to yield a rough count of words that have been "learned", at least partially, during the course of an intervention.

However, there are several drawbacks to PPVT-like measures' inability to tap into deeper semantic knowledge when used in the context of a shorter-term vocabulary intervention. The first drawback is that such measures cannot distinguish between low and high quality lexical representations. For example, if a child is able to recognize an image of a *shield* at the beginning of an intervention, she would probably be assessed as "knowing" that word on a PPVT-like pretest. If after the intervention she can not only recognize *shield* but can also say that it is used by knights to protect themselves, she will still simply get a single point on a PPVT-like posttest. As measured by the assessment, her knowledge of the word *shield* has remained static.

PPVT-like measures' inability to tap into deeper knowledge is also problematic when comparing instructional methods. For example, a book-reading and play vocabulary intervention for preschoolers compared children's learning of words that were explicitly taught ("target words") to words that were used in the book but not explained ("exposure words") (Dickinson et al., 2017). Children's learning of target and exposure words was measured on both a PPVT-like measure and a depth measure. There was no significant difference in children's learning of target and exposure words on the PPVT-like measure, but there was a significant difference and a medium effect size

($d = 0.50$) between the learning of target and exposure words on the depth measure. If only the PPVT-like measure had been used, it would appear that there was no difference between children's learning of words they were merely exposed to vs words they were explicitly taught.

PPVT-like measures can also be limited in the types of semantic knowledge they tap into simply because all words tested must be visually portrayed. It is much more feasible for researchers to create test items for words that are highly concrete and imageable, such as concrete nouns and concrete verbs. This is problematic because we know that children need to learn words from a variety of form classes (Harris et al., 2011) and words that will be helpful across a variety of academic domains (Beck et al., 2013). Such words are often quite abstract and difficult to depict. For example, *contradict*, *precede* and *auspicious* (Beck et al., 2013) are all Tier Two words, none of which suggest a clear, representative image that could be used without confusion on a pictorial assessment for young children.

Use. Neither the PPVT or PPVT-like measures assess for knowledge of a word's use in context. Each word is read aloud by an examiner without additional comment.

The Expressive Vocabulary Test (EVT), the Expressive One-Word Picture Vocabulary Test (EOWPVT) and targeted measures modelled after each. The EVT (Williams, 2007) and EOWPVT (Brownell, 2010) are vocabulary breadth measures in which children are asked to name an item depicted by a single picture. Researchers have also designed targeted assessments modelled after the EVT and EOWPVT, in which children are shown a picture of a target word and asked to name the picture (e.g. Blewitt et al., 2009; Hargrave and Sénéchal, 2000; Whitehurst et al., 1994).

Form, meaning and use. This type of assessment is more demanding than the PPVT or PPVT-like measures, because children must produce a correct phonological representation for a word. Therefore, both a limited understanding of form and meaning, although not use, are tapped into by this measure.

Woodcock-Johnson III Picture Vocabulary Subtest (WJ-III). The WJ-III (Woodcock et al., 2001) is a less frequently used standardized breadth measure that includes both receptive and expressive picture vocabulary items. Its ability to measure form, meaning and use is similar to the PPVT-IV and EVT/EOWPVT.

Depth measures

There is no single approach for assessing depth that dominates the field. Including a depth measure in a vocabulary intervention or correlational study is by no means common (e.g. in the 2010 meta-analysis of vocabulary interventions by Marulis and Neuman, nearly three-quarters of studies did not use a depth measure), so the field has yet to reach a consensus as to which depth measures are most valid and theoretically sound. We review here depth assessments that have been used with young children aged preK-1st grade in vocabulary interventions or studies used to predict reading comprehension from oral language skills.

Definition tasks. Definition tasks have been used frequently as depth measures, both in standardized and targeted forms. Children are typically asked what they know about a word, and their responses can either be scored (1) along a continuum, so that fewer points are given for more connotative or contextual responses and more points for decontextualized responses (Biemiller and Boote, 2006; Coyne et al., 2009; Leung, 2008), or (2) for completeness of definition, in which a point is earned for each unit of semantic information given (Blewitt et al., 2009). For example, the Blewitt et al. (2009) study awarded a point for each of the following types of semantic information: superordinate category membership, synonyms, perceptual or functional properties, and parts. In other studies, points are also awarded for using the target word in a typical context or for representing the word with a gesture (Hadley et al., 2016).

Standardized definition tasks such as the Oral Vocabulary subtest of the Test of Language Development-Primary (TOLD:P-4, now in version 4, Hammill and Newcomer, 2008) have been used to track the relationship between early oral language and later reading ability (Roth et al., 2002). Similarly, the Word Definitions subtest of the Test of Word Knowledge (TOWK, Wiig and Secord, 1992) has been used to assess the relationships between vocabulary, word-reading, and reading comprehension (Ouellette, 2006). Targeted definition tasks have been used to track increases in children's knowledge of a specific set of words taught during an intervention (Biemiller and Boote, 2006; Blewitt et al., 2009; Coyne et al., 2009; Hadley et al., 2016; Leung, 2008).

Form. At a minimum, definition tasks require children to recognize a word's phonological representation and retrieve its associated semantic information. Often, children produce the word when giving a definition, and/or provide

inflected versions of the word, but this information is not explicitly scored for in the definition tasks listed above.

Meaning. Definition tasks are highly demanding and tap into knowledge of a word at the high end of the lexical quality continuum, as they require children to provide a variety of information about word meaning, such as what something does or is used for (functional knowledge), or to make connections to broader conceptual knowledge, such as category membership. Moreover, they tap metalinguistic skills as they require children to think about their lexicon and express their knowledge of word meaning explicitly (Roth et al., 2002). Children receive more points for definitions that include more (Blewitt et al., 2009; Hadley et al., 2016) or higher-level (e.g. Coyne et al., 2009) semantic information, allowing for differentiation between low and high lexical quality representations.

Use. Some definition tasks code strictly for decontextualized semantic information (Blewitt et al., 2009). Others give credit for definitions that provide an example of a word's use, or even a typical association with a particular context (Hadley et al., 2016). We argue that such scoring is appropriate for young children since they have little experience with formal definitions (Snow et al., 1991) and often demonstrate their knowledge of a word by giving an example in context (Hadley et al., 2016), therefore showing their ability to use a word in a real-world setting.

One drawback of definition tasks is that they require a great deal of oral language proficiency and children do not, or cannot, always express all of the semantic knowledge that they have about a word. While administration is easy and the format of the test is highly flexible and requires little advance preparation, these tasks usually require coding of children's responses (but note that the TOWK and TOLD:P-4 can be scored on the spot).

Clinical Evaluation of Language Fundamentals (CELF), Formulated Sentences subtest. For the Formulated Sentences subset of the CELF (Semel et al., 2003), children are shown a picture and asked to generate a sentence describing it while using a target word. Scoring ranges from 0 to 2 and responses are scored for whether the use of the target word is syntactically, pragmatically and semantically correct. This assessment has been used as a measure of depth for its ability to capture multiple dimensions of word knowledge in a study exploring the relationships between depth, breadth and reading comprehension (Proctor et al., 2012), but its use for this purpose is not widespread.

Form, meaning, and use. The Formulated Sentences subtest taps into children's knowledge of form (particularly grammatical aspects), conceptual knowledge and correct usage. This measure is unique in its ability to quickly assess all three aspects of depth, and highly promising for the same reason, although it has not been widely used outside clinical settings (but see Proctor et al., 2012).

Closed-ended questions. In this method, children are asked several questions about each target word (e.g. Beck and McKeown, 2007; Coyne et al., 2009; Kearns and Biemiller, 2010). Some closed-ended question measures ask about both the meaning and use in a typical context for a word (e.g. Beck and McKeown, 2007). For example, such questions might ask: ''Does *extraordinary* mean very hungry?'' and ''Would it be *extraordinary* to see a monkey at school?'' (Beck and McKeown, 2007). Others ask only about a word's use in context (Kearns and Biemiller, 2010). Children are typically assessed as ''knowing'' a word if they answer the majority of questions for that word correctly (e.g. McKeown and Beck, 2014).

Form, meaning and use. This measure assesses children's conceptual knowledge of words as well as their ability to understand typical contexts of use. Children are not required to produce the word, its meaning or use it in a sentence, but they must retrieve a representation based on its phonological representation and recognize correct use(s) and correct meaning(s) in order to earn full points. Such a task is significantly more demanding than breadth measures such as the PPVT as it taps into children's semantic and contextual knowledge of target words, but less demanding than definition tasks that ask children to produce semantic information. Note also that the demands of this task depend greatly on the difficulty of the questions asked: Coyne et al. (2009) included two types of closed-ended question tasks, one which tapped into high-quality knowledge, and one that tapped into partial knowledge. Administration and scoring of this task is quick and straightforward and can be done in groups even for children as young as kindergarten (Beck and McKeown, 2007; Kearns and Biemiller, 2010).

Story-retelling and picture description measures. Story retelling tasks, in which children's use of target vocabulary words is evaluated for accuracy as they retell the narrative of a book, have also been used as a measure of high quality word knowledge (Penno et al., 2002) because they reveal students' ability to retrieve, pronounce and use a word correctly in context. This measure is

somewhat problematic in that children may know, but choose not to use, the target vocabulary. It is also unclear from this measure whether children's knowledge of a word has generalized beyond the storybook context. Moreover, the scoring of this task is more difficult, in that it requires transcription and coding.

A more focused version of a story-retelling measure has been used by McKeown and Beck (2014), in which students are shown a picture for each target word (e.g. two girls with their arms around each other for the word *inseparable*). Children are asked specific questions about the picture (e.g. "What can you see by looking at the girls?"). This version of a storytelling measure taps into higher-quality knowledge of use because it asks children to transfer their knowledge of a word to a new context, but it is subject to the same problem of children potentially knowing, but choosing not to use, the target word.

Form, meaning and use. Both of these versions of story-telling measures are quite demanding because they test children's ability to retrieve an appropriate target word and use it correctly in context, a skill that demonstrates a high quality, coherent lexical representation with well-established semantic, syntactic and phonological information for words.

Context integration tasks. McKeown and Beck (2014) used a highly demanding vocabulary measure called a context integration task that tests the quality of kindergarten children's lexical representations at the higher end of the lexical continuum. Children are asked a question that probes for understanding of a target word in context. For example, the following question was asked for the word *insist*: "Jim had to *insist* that Freddy go on the merry-go-round. How did Freddy feel about the merry-go-round?" Children's responses were given 1 point if they reflected knowledge of the target word (e.g. "he didn't like it") and 0 points if they were incorrect (e.g. "it was fun").

Form, meaning and use. This measure is especially demanding because the sentence context encourages children to interpret the target word incorrectly (i.e. in the example given above, the positive connotation of "merry-go-round" conflicts with the negative connotation of "insist"). Therefore, the child must have very high quality semantic and pragmatic knowledge of a word in order to "crowd out" this alternative suggestion. This measure also does not provide any pictorial prompts to help support children's knowledge of target words.

One possible issue with this measure is that it requires strong listening comprehension and self-regulation. McKeown and Beck (2014) used this measure with kindergarten and 1st-grade children, but it may be too demanding for preschool children. This measure is relatively easy to administer and allows for on-the-spot scoring.

Conclusions and recommendations

This paper began by citing Pearson et al.'s call for more conceptually rich measures in the field of vocabulary assessment. In looking more specifically at vocabulary measures intended for the early childhood population, it is clear that the field is indeed, as both Pearson et al. (2007) and Hoffman et al. (2014) conclude, "undernourished". One reason for this undernourishment is that many measures are not grounded in a theoretical perspective on the word-learning process. Here, we have used Perfetti's LQH (2007) as our primary conceptual framework to explore how commonly used measures assess different aspects of form, meaning and use, as well as varying levels of lexical quality. By applying this framework, we have made visible (see Table 1) the inability of many widely used vocabulary measures to tap into multiple facets of word knowledge. We review our main findings and provide some more specific recommendations below.

Breadth measures: Strengths and limitations

Breadth measures are currently the most widely used type of measure with early childhood learners (e.g. Marulis and Neuman, 2010; NELP, 2008). Breadth measures are useful when employed for certain purposes: standardized breadth measures such as the PPVT are invaluable for diagnostic assessment and can be used to track long-term growth in vocabulary knowledge. Vocabulary breadth has also been shown to predict later reading comprehension in correlational studies (e.g. Cunningham and Stanovich, 1997). Finally, targeted breadth measures can serve as a quick, although not particularly sensitive, way to gauge children's growth in word knowledge.

However, as Table 1 illustrates, breadth measures focus on the recognition of phonological information and perceptual features, but they do not tap into higher-level semantic knowledge or knowledge of use. In other words, an overreliance on breadth measures means that we do not know if children can use or understand words in ways that, according to Perfetti's LQH (2007), are important for listening or reading comprehension.

The field's emphasis on breadth also has consequences for instruction: there is a large number of intervention studies that have successfully fostered receptive, broad vocabulary knowledge (e.g. Hargrave and Sénéchal, 2000), but the research base supporting high-quality, flexible knowledge is more limited (but see McKeown and Beck, 2014). This gap in the field is especially consequential because a growing body of research has suggested that if vocabulary instruction is to impact on reading comprehension, efforts aimed at increasing vocabulary size alone may not be sufficient: learners must also gain deep, flexible knowledge of words so that meanings can be quickly accessed when needed (Beck and McKeown, 2007; Silverman and Hartranft, 2015; Stahl and Fairbanks, 1986).

Depth measures: Strengths and limitations

We argue here that much can be learned from shifting our focus to a multifaceted strategy for assessing word knowledge, one that includes both breadth and depth. Depth measures allow for information to be gathered about the quality of children's phonological, syntactic, semantic, and pragmatic representations of words (see Table 1). The value of these measures has been shown both in correlational work (e.g. depth was more strongly predictive of reading comprehension than breadth in the NELP meta-analysis, 2008), and in interventions in which depth measures have proved to be more sensitive (e.g. Coyne et al., 2009; Dickinson et al., 2017) and provide more detailed information about word-learning (Hadley et al., 2016) than breadth measures. However, depth measures can be quite time-consuming to administer and score, and their lack of widespread use means that the field is still coming to a consensus on which measures have the most predictive power and can best inform instruction.

Recommendation: Using a battery of measures

One promising response to Pearson et al.'s (2007) call for "conceptual richness" in the field of vocabulary assessment is the use of multiple measures to assess children's word learning. However, these multiple measures must provide *different* kinds of information about learning: some intervention studies use both a receptive and an expressive pictorial assessment, but as both of these measures only tap into shallow lexical representations, the results may be redundant and represent a missed opportunity to test for higher quality lexical representations.

The benefit of using multiple vocabulary measures that tap into knowledge along the lexical quality continuum can be illustrated by a study reported by Coyne et al. (2009) in which two instructional conditions are compared, one ("Rich Instruction") more intensive than the other ("Embedded Instruction"). This study found that the additional support provided by the Rich Instruction condition led to more refined word knowledge, while the Embedded Instruction condition built only partial knowledge of words. This finding was visible only because a battery of four measures, ranging in difficulty, was used to gauge learning. Researchers used two measures to tap into low-quality representations: (1) a closed-ended question measure that asked low-level questions about the target word; and (2) a yes/no task where children had to say whether a definition of word was correct or incorrect; and two measures to tap higher-quality knowledge: (1) a definition task; and (2) a closed-ended question measure that asked high-level questions about words' use in context. Interestingly, the measures tapping into low lexical quality showed no difference in learning between the two conditions: only the measures tapping into high lexical quality showed that Rich Instruction was more effective.

This study, along with McKeown and Beck (2014), who used a similar battery of assessments, represents important steps in a promising direction. Both studies articulate a theoretically grounded view of vocabulary knowledge, set out clear goals for learning, and use multiple measures that tap into learning at various points along the lexical quality continuum. For example, McKeown and Beck (2014) use a cognitive processing framework to inform their methods and measures, and they use measures that range in difficulty to determine whether children have fully integrated multiple aspects of word knowledge for learned words.

Based on this work and our theoretical framework, we suggest the use of a battery of targeted measures for vocabulary interventions: (1) a targeted breadth measure that requires only surface-level knowledge and allows for a rough count of words "learned" to capture minimal levels of knowledge (a receptive measure such as a carefully designed pictorial assessment); (2) a targeted depth measure that emphasizes children's ability to *use* a word accurately (such as a researcher-created version of the CELF-Formulated Sentences task, Semel et al., 2003); and (3) a depth measure that tests for very high quality knowledge of words and emphasizes knowledge of *meaning* (such as a definition task, Hadley et al., 2016). This collection of measures ensures that low levels of knowledge are captured and that relatively subtle differences in learning can be pinpointed by highly sensitive depth measures.

We also suggest the use of the PPVT or other standardized measure at pretest to gauge the size of children's vocabularies in relation to their peers and to allow for comparability of samples across studies.

Recommendation: Further development and use of standardized depth measures

Standardized depth measures have been shown to be of value in predicting later reading comprehension (National Early Literacy Panel, 2008), but the use of such measures is not yet widespread. Future research should explore how the inclusion of general depth measures can add to our understanding of the relationship between reading comprehension and vocabulary. It would be helpful to determine whether one or two aspects of depth could be tested that demonstrate knowledge of the others – for example, if a child can use a word in several contexts correctly, can we then also assume that their semantic knowledge is well-developed? Such information would allow us to better streamline assessment.

It would also be helpful to develop a depth assessment that tested children's semantic knowledge of the relationships between words. These types of depth measures exist for older children, such as semantic association tasks in which students must select the words that are related to a target word from among several choices (e.g. the Word Association Test, validated for ages 9–12, Schoonen and Verhallen, 2008), and could be adapted for use with younger children. For example, a semantic association task for preschool and kindergarten-aged children might ask them to sort several pictures of words or objects into groups that "go together" in order to assess their knowledge of the relationships between these words. This type of measure would be a potentially fruitful addition to the standardized vocabulary measures currently available.

Recommendation: New ways to test for generalization

We also hope that future research explores the use of other measures, standardized or researcher-created, that test for generalization of knowledge. While vocabulary interventions should measure the learning of taught words, the ultimate goal is to improve children's general vocabularies and perhaps even word-learning abilities. Innovative assessments that can detect changes in vocabulary abilities are needed. For example, assessments such as the Diagnostic Evaluation of Language Variation (DELV, Seymour et al., 2005)

and the Quick Interactive Language Screener (QUILS, Golinkoff et al., 2017) contain tasks that could be adapted to test for changes in children's word-learning ability, such as the DELV Fast Mapping test, which tests a child's ability to infer the meaning of a new word after hearing it in context. Similarly, Neuman et al. (2011) developed a task that tested children's ability to categorize new words after participating in an intervention in which they were taught words in taxonomic categories. More widespread use of similar assessments could provide a novel view into not only whether vocabulary interventions can change what children *know*, but whether they can change how they *learn*.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education (grant R305A110128).

References

Ainsworth S, Welbourne S and Hesketh A (2015) Lexical restructuring in preliterate children: Evidence from novel measures of phonological representation. *Applied Psycholinguistics*, 37(4), 997–1023.

Anderson RC and Freebody P (1981) *Vocabulary Knowledge*. Newark: International Reading Association.

Beck IL and McKeown MG (2007) Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal* 107(3): 251–271.

Beck IL, McKeown MG and Kucan L (2013) *Bringing Words to Life: Robust Vocabulary Instruction*. New York: The Guilford Press.

Biemiller A and Boote C (2006) An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology* 98(1): 44–62.

Bion RAH, Borovsky A and Fernald A (2013) Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition* 126(1): 39–53.

Blewitt P, Rump KM, Shealy SE, et al. (2009) Shared book reading: When and how questions affect young children's word learning. *Journal of Education Psychology* 101(2): 294–304.

Bolger DJ, Balass M, Landen E, et al. (2008) Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes* 45(2): 122–159.

Booth AE and Waxman SR (2002) Word learning is 'smart': Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition* 84(1): B11–B22.

Borovsky A, Ellis EM, Evans JL, et al. (2016) Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science* 19(6): 918–932.

Brownell R (2010) *Expressive One-Word Picture Vocabulary Test, Fourth Edition (EOWPVT)*. Novato, CA: Academic Therapy Publications.

Cain K and Oakhill JV (2014) Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique* 114: 647–662.

Carey S (1978) The child as word learner. In: Halle M, Bresnan J and Miller A (eds) *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press, pp. 264–293.

Clark EV (2010) *First Language Acquisition*, 2nd ed. Cambridge: Cambridge University Press.

Coyne MD, McCoach DB, Loftus S, et al. (2009) Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal* 110(1): 1–18.

Cunningham AE and Stanovich KE (1997) Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology* 33(6): 934–945.

Dickinson DK, Golinkoff RM and Hirsh-Pasek K (2010) Speaking out for language: Why language is central to reading development. *Educational Researcher* 39(4): 305–310.

Dickinson DK, Nesbitt KT, Collins MF, et al. (2017) *Using Book-Reading to Teach for Breadth and Depth*. Epub ahead of print.

Dickinson DK and Porche MV (2011) Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development* 82(3): 870–886.

Dunn LM and Dunn LM (2007) *Peabody Picture Vocabulary Test, IV*. Circle Pines, MN: American Guidance Service, Inc.

Estes KG, Evans JL, Alibali MW, et al. (2007) Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science* 18(3): 254–260.

Farkas G and Beron K (2004) The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research* 33(3): 464–497.

Forbes JN and Farrar MJ (1993) Children's initial assumptions about the meaning of novel motion verbs: Biased and conservative? *Cognitive Development* 8(3): 273–290.

Frishkoff GA, Perfetti CA and Collins-Thompson K (2011) Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading* 15(1): 71–91.

Gentner D (2006) Why verbs are hard to learn. In: Hirsh-Pasek K and Golinkoff RM (eds) *Action Meets Word: How Children Learn Verbs*. Cambridge: Oxford University Press.

Golinkoff RM, de Villiers J, Hirsh-Pasek K, et al. (2017) *User's Manual for the Quick Interactive Language Screener (QUILS): A Measure of Vocabulary, Syntax, and Language Acquisition Skills in Young Children*. Baltimore: Brookes Publishing.

Hadley EB, Dickinson DK, Hirsh-Pasek K, et al. (2016) Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly* 51(2): 181–198.

Hammill D and Newcomer P (2008) *Test of Language Development-Primary, fourth edition (TOLD-P:4)*. Austin, TX: Pro-Ed.

Hargrave AC and Sénéchal M (2000) A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly* 15(1): 75–90.

Harris J, Golinkoff RM and Hirsh-Pasek K (2011) Lessons from the crib for the classroom: How children really learn vocabulary. In: Neuman SB and Dickinson DK (eds) *Handbook of Early Literacy Research*. New York: Guilford Press, pp. 49–66.

Hoffman JL, Teale WH and Paciga KA (2014) Assessing vocabulary learning in early childhood. *Journal of Early Childhood Literacy* 14(4): 482–509.

Hollich G, Hirsh-Pasek K and Golinkoff RM (2000) II. The Emergentist Coalition Model. *Monographs of the Society for Research in Child Development* 65(3): 17–29.

Horst JS and Samuelson LK (2008) Fast mapping but poor retention by 24-month-old infants. *Infancy* 13(2): 128–157.

Kearns G, & Biemiller A. (2010). Two questions vocabulary assessment: Developing a new method for group testing in kindergarten through second grade. *The Journal of Education*, 190(1/2), 31–41.

Kennedy MM (2007) Defining a literature. *Educational Researcher* 36(3): 139–147.

Kieffer MJ and Lesaux NK (2012) Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing* 25(2): 347–373.

Kucan L and Palincsar AS (2010) Locating struggling readers in a reconfigured landscape. In: Kamil M, Pearson PD, Moje EB, et al. (eds) *Handbook of Reading Research*. Vol. 4, New York: Routledge. p.341.

Lahey M (1988) *Language Disorders and Language Development*. Needham, MA: Macmillan.

Lee J (2011) Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics* 32(01): 69–92.

Leung CB (2008) Preschoolers' acquisition of scientific vocabulary through repeated read-aloud events, retellings, and hands-on science activities. *Reading Psychology* 2928.

McKeown MG and Beck IL (2014) Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly* 29(4): 520–530.

Marulis LM and Neuman SB (2010) The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research* 80(3): 300–335.

Melka F (1997) Receptive versus productive aspects of vocabulary. In: Schmitt N and McCarthy M (eds) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 84–102.

Miller GA and Fellbaum C (1991) Semantic networks of English. *Cognition* 41(1–3): 197–229.

Nagy WE and Scott JA (2000) Vocabulary processes. In: Kamil M, Mosenthal P, Pearson PD, et al. (eds) *Handbook of Reading Research*. Vol. 3, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 269–284.

Nation K (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

National Early Literacy Panel (2008) *Developing Early Literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

National Reading Panel. (2000) *Report of the National Reading Panel: Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. Washington, DC: US Government Printing Office.

Neuman SB, Newman EH and Dwyer J (2011) Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly* 46(3): 249–272.

Ordóñez CL, Carlo MS, Snow CE, et al. (2002) Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology* 94(4): 719–728.

Ouellette GP (2006) What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology* 98(3): 554–566.

Pearson PD, Hiebert EH and Kamil ML (2007) Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly* 42(2): 282–296.

Penno JF, Wilkinson IAG and Moore DW (2002) Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew Effect? *Journal of Educational Psychology* 94(1): 23–33.

Perfetti C (2007) Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading* 11(4): 357–383.

Perfetti C (2013) Representations and awareness in the acquisition of reading competence. In: (Rieben L and Perfetti C (eds) *Learning to Read: Basic Research and its Implications*. New York: Routledge, pp.33–47.

Pollard-Durodola SD, Gonzalez JE, Simmons DC, et al. (2011) The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children* 77(2): 161–183.

Proctor CP, Silverman RD, Harring JR, et al. (2012) The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing* 25(7): 1635–1664.

Qi C, Kaiser AP, Milan S, et al. (2006) Language performance of low-income African-American and European American preschool children on the PPVT-III. *Language, Speech & Hearing Services in Schools* 37(1): 5–16.

Read J (2004) Plumbing the depths: How should the construct of vocabulary knowledge be defined? In: Bogaards P, Laufer B, et al. (eds) *Vocabulary in a Second Language:*

Selection, Acquisition, and Testing. Philadelphia, PA: John Benjamins Publishing Company, pp. 209–227.

Roskos K, Ergul C, Bryan T, et al. (2008) Who's learning what words and how fast? Preschoolers' vocabulary growth in an early literacy program. *Journal of Research in Childhood Education* 22(3): 275–290.

Roth FP, Speece DL and Cooper DH (2002) A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research* 95(5): 259–272.

Schmitt N (2014) Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64(4): 1–39. DOI:10.1111/lang.12077.

Schoonen R and Verhallen M (2008) The assessment of deep word knowledge in young first and second language learners. *Language Testing* 25(2): 211–236.

Semel E, Wiig EH and Secord W (2003) *Clinical Evaluation of Language Fundamentals – Fourth Edition*. London: Pearson.

Senechal M (1997) The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language* 24(1): 123–138.

Senechal M, Ouellette G and Rodney D (2006) The misunderstood giant: On the predictive role of early vocabulary to future reading. In: Dickinson DK and Neuman SB (eds) *Handbook of Early Literacy Research*, vol. 2. New York: Guilford Press, pp.173–182.

Seston R, Golinkoff RM, Ma W, et al. (2009) Vacuuming with my mouth? Children's ability to comprehend novel extensions of familiar verbs. *Cognitive Development* 24(2): 113–124.

Seymour HN, Roeper TW and de Villiers J (2005) *Diagnostic Evaluation of Language Variation—Norm Referenced (DELV)*. London: Pearson.

Silverman RD and Crandell JD (2010) Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly* 45(3): 318–340.

Silverman RD and Hartranft AM (2015) *Developing Vocabulary and Oral Language in Young Children*. New York: Guilford Press.

Snow CE, Cancino H, De Temple J, et al. (1991) Giving formal definitions: A linguistic or metalinguistic skill? In: Bialystock E (ed.) *Language Processing in Bilingual Children*. Cambridge: Cambridge University Press, pp.90–112.

Stahl SA and Fairbanks MM (1986) The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research* 56(1): 72–110.

Swingley D (2007) Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology* 43(2): 454–464.

Swingley D (2010) Fast mapping and slow mapping in children's word learning. *Language learning and Development* 6(3): 179–183.

Tannenbaum KR, Torgesen JK and Wagner RK (2006) Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading* 10(4): 381–398.

Verhoeven L and Perfetti CA (2011) Introduction to this special issue: Vocabulary growth and reading skill. *Scientific Studies of Reading* 15(1): 1–7.

Wasik BA and Bond MA (2001) Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Education Psychology* 93(2): 243–250.

Whitehurst GJ, Arnold DS, Epstein JN, et al. (1994) A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology* 30(5): 679–689.

Wiig EH and Secord W (1992) *Test of Word Knowledge*. San Antonio, TX: Psychological Corporation.

Williams K (2007) *Expressive Vocabulary Test, second edition (EVT)*. London: Pearson.

Woodcock R, McGrew KS and Mather N (2001) *Woodcock Johnson III Test of Cognitive Abilities*. Rolling Meadows, IL: Riverside Publishing.

Yu C and Smith LB (2012) Modeling cross-situational word–referent learning: Prior questions. *Psychological Review* 119(1): 21–39.

Yuan S and Fisher C (2009) "Really? She blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science* 20(5): 619–626.

Yurovsky D, Fricker DC, Yu C, et al. (2014) The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review* 21(1): 1–22.

Zimmerman IL, Steiner VG and Pond RE (2002) *Preschool Language Scale, Fourth Edition (PLS-4)*. London: Pearson.